

WHAT IS CLAIMED IS:

1. A speech signal processing apparatus comprising:
speech segment search means for searching a
speech database for speech segments that satisfy a
5 phonetic environment;

HMM learning means for computing HMMs of phonemes
on the basis of a search result of said speech segment
search means;

- segment recognition means for performing segment
10 recognition of the speech segments on the basis of the
HMMs of the phonemes; and

registration segment determination means for
determining a speech segment to be registered in a
segment dictionary in accordance with a segment
15 recognition result of said segment recognition means.

2. The apparatus according to claim 1, wherein said
segment recognition means adopts diphones as units of
the phonemes, categorizes speech segments into four
categories CC, CV, VC, and VV (C: a consonant, V: a
20 vowel), and performs segment recognition in each
category.

3. The apparatus according to claim 1, wherein said
registration segment determination means comprises:

pattern storage means which registers allowable
25 speech segment patterns, and

said registration segment determination means
checks if a speech segment pattern which matches a

speech segment that is not successfully recognized by said segment recognition means, and registers that speech segment in the segment dictionary if the corresponding speech segment pattern is found.

5 4. The apparatus according to claim 1, wherein said registration segment determination means registers a speech segment in the segment dictionary when the number of speech segments recognized by said speech segment recognition means is not less than a
10 predetermined value.

5. The apparatus according to claim 4, wherein said registration segment determination means registers a speech segment in the segment dictionary if at least a vowel part of the speech segment is correctly
15 recognized, even when the number of speech segments recognized by said speech segment recognition means is not more than a predetermined value.

6. The apparatus according to claim 1, wherein said segment recognition means computes likelihoods of
20 speech segments of an identical phoneme, and
said registration segment determination means registers, in the segment dictionary, speech segments having upper likelihoods or having likelihoods not less than a predetermined value.

25 7. The apparatus according to claim 6, wherein said registration segment determination means registers, in the segment dictionary, speech segments having upper

values obtained by normalizing the likelihoods by durations of the speech segments or likelihoods having the values not less than a predetermined value.

8. A speech signal processing method comprising:

5 the speech segment search step of searching a speech database for speech segments that satisfy a phonetic environment;

the HMM learning step of computing HMMs of phonemes on the basis of a search result of the speech
10 segment search step;

the segment recognition step of performing segment recognition of the speech segments on the basis of the HMMs of the phonemes; and

the registration segment determination step of
15 determining a speech segment to be registered in a segment dictionary in accordance with a segment recognition result of the segment recognition step.

9. The method according to claim 8, wherein the segment recognition step adopts diphones as units of
20 the phonemes and categorizes speech segments into four categories CC, CV, VC, and VV (C: a consonant, V: a vowel), and includes the step of performing segment recognition in each category.

10. The method according to claim 8, wherein the
25 registration segment determination step comprises:

the pattern storage step of registering allowable speech segment patterns, and

the registration segment determination step includes the step of checking if a speech segment pattern which matches a speech segment that is not successfully recognized in the segment recognition step,
5 and registering that speech segment in the segment dictionary if the corresponding speech segment pattern is found.

11. The method according to claim 8, wherein the registration segment determination step includes the
10 step of registering a speech segment in the segment dictionary when the number of speech segments recognized in the speech segment recognition step is not less than a predetermined value.

12. The method according to claim 11, wherein the
15 registration segment determination step includes the step of registering a speech segment in the segment dictionary if at least a vowel part of the speech segment is correctly recognized, even when the number of speech segments recognized in the speech segment
20 recognition step is not more than a predetermined value.

13. The method according to claim 8, wherein the segment recognition step includes the step of computing likelihoods of speech segments of an identical phoneme,
and

25 the registration segment determination step includes the step of registering, in the segment

dictionary, speech segments having upper likelihoods or having likelihoods not less than a predetermined value.

14. The method according to claim 13, wherein the registration segment determination step includes the
5 step of registering, in the segment dictionary, speech segments having upper values obtained by normalizing the likelihoods by durations of the speech segments or likelihoods having the values not less than a predetermined value.

10 15. A computer readable storage medium storing a program for implementing a method cited in claim 8.

16. A speech signal processing apparatus comprising:
a segment dictionary in which speech segments are registered by a method cited in claim 8;

15 language analysis means for performing language analysis of input text data;

prosody generation means for generating prosody on the basis of an analysis result of said language analysis means;

20 speech segment selection means for search said segment dictionary on the basis of the prosody generated by said prosody generation means to select corresponding speech segments;

speech segment modification/concatenation means
25 for modifying and concatenating the speech segments selected by said speech segment selection means; and

speech reproduction means for reproducing speech on the basis of the result modified by said speech segment modification/concatenation means.

17. A speech signal processing apparatus comprising:

5 HMM learning means for learning HMMs corresponding to phonemes using a plurality of speech segments that satisfy a predetermined phonetic environment; and

registration segment determination means for selecting a speech segment to be registered in a
10 segment dictionary used in speech synthesis on the basis of the HMMs corresponding to the phonemes.

18. The apparatus according to claim 17, wherein said registration segment determination means obtains a maximum likelihood HMM which has a maximum likelihood
15 with one of the plurality of speech segments from the HMMs corresponding to the phonemes, checks if the one speech segment is a speech segment used in learning of the maximum likelihood HMM, and selects the one speech segment when the one speech segment is a speech segment
20 used in learning of the maximum likelihood HMM.

19. The apparatus according to claim 17, further comprising speech synthesis means for producing synthetic speech using the segment dictionary.

20. A speech signal processing method comprising:
25 the HMM learning step of learning HMMs corresponding to phonemes using a plurality of speech

segments that satisfy a predetermined phonetic environment; and

the registration segment determination step of selecting a speech segment to be registered in a
5 segment dictionary used in speech synthesis on the basis of the HMMs corresponding to the phonemes.

21. The method according to claim 20, wherein the registration segment determination step includes the step of obtaining a maximum likelihood HMM which has a
10 maximum likelihood with one of the plurality of speech segments from the HMMs corresponding to the phonemes, checking if the one speech segment is a speech segment used in learning of the maximum likelihood HMM, and selecting the one speech segment when the one speech
15 segment is a speech segment used in learning of the maximum likelihood HMM.

22. The method according to claim 20, further comprising the speech synthesis step of producing synthetic speech using the segment dictionary.

20 23. A computer readable program storing a program for implementing a method cited in claim 20.